Magdalena PIEKUTOWSKA[1], Gniewko NIEDBAŁA[2] Mariusz ADAMSKI[2], Mirosław CZECHLOWSKI[2], Tomasz WOJCIECHOWSKI[2], Aneta CZECHOWSKA-KOSACKA[3], Katarzyna WÓJCIK OLIVEIRA[3]
[1] Koszalin University of Technology, Department of Agrobiotechnology, Faculty of Mechanical Engineering
ul. Racławicka 15-17, 75-620 Koszalin, Poland
[2] Poznań University of Life Sciences, Institute of Biosystems Engineering, Faculty of Agronomy and Bioengineering
ul. Wojska Polskiego 50, 60-627 Poznań, Poland
[3] Lublin University of Technology Institute of Environmental Protection Engineering, Faculty of Environmental Engineering
ul. Nadbystrzycka 40B, 20-618 Lublin, Poland
e-mail: magdalena.piekutowska@poczta.fm; gniewko@up.poznan.pl; k.wojcik-oliveira@pollub.pl

Received: 2018-09-19 ; Accepted: 2018-12-04

MODELING METHODS OF PREDICTING POTATO YIELD - EXAMPLES AND POSSIBILITIES OF APPLICATION

Summary

The purpose of the following work is to review the methods used in predicting plant yields, with particular emphasis on potato production. The article refers to the histological methods of estimating plant yields and prevailing trends: ground-based remote sensing, which is often associated with regression calculus, multiple regression, artificial intelligence and image analysis. There are also two popular models SUBSTOR and LINTUL-POTATO, which are the foundation for developing more and more accurate tools of potato yield estimation. There are many methods that allow to predict yields before the end of the growing season. The most important element in creating prediction models is choosing the appropriate number of independent variables that actually shape the yielding of potatoes. Timely and accurate prediction of crop yields improve the management of agricultural production as well as limit financial, quantitative and qualitative losses of crops.
***Key words**: yield prediction, potato, artificial neural networks, regression*

METODY MODELOWANIA PREDYKCJI PLONU ZIEMNIAKÓW – PRZYKŁADY I MOŻLIWOŚCI ZASTOSOWANIA

Streszczenie

Celem niniejszej pracy był przegląd metod wykorzystywanych w prognozowaniu plonów roślin ze szczególnym uwzględnieniem produkcji ziemniaka. W artykule nawiązano do historycznych sposobów szacowania plonów roślin oraz obecnie panujących trendów w predykcji: teledetekcji naziemnej, która często powiązana jest z rachunkiem regresyjnym, regresji wielorakiej, sztucznej inteligencji, analizie obrazów. Wspomniano także o dwóch popularnych modelach SUBSTOR i LINTUL-POTATO, które stworzyły podwaliny do opracowywania coraz dokładniejszych narzędzi prognozujących plony ziemniaków. Wiele metod pozwala na predykcję plonów przed zakończeniem sezonu wegetacyjnego. Najistotniejszym elementem tworzenia modeli predykcyjnych jest dobór odpowiedniej liczby zmiennych niezależnych, które rzeczywiście kształtują plonowanie ziemniaków. Terminowe i dokładne prognozy plonów roślin uprawnych usprawniają zarządzanie produkcją rolniczą, pozwalają na ograniczanie strat finansowych, ilościowych i jakościowych plonów.
***Słowa kluczowe**: predykcja plonowania, ziemniaki, sztuczne sieci neuronowe, regresja*

1. Introduction

Plant yield prediction deals with estimation of the yield value of a given crop species, assuming that the conditions prevailing during the growing season were close to those prevailing many years before, in short, it's a scientific way to predict future events. Forecasting made before the harvest influences the organization of agricultural production, plays a significant role in the functioning of domestic and foreign food markets [26]. Potato is a species that requires large capital expenditure during cultivation. Farmers who want to expand potato production areas should be able to determine the yield in advance that can be achieved in a particular season, so as to consider the scale of costs being incurred (profitability of production) [7] and individual storage capabilities. The growing interest in potato cultivation globally in connection with the decreasing area of land in agricultural use introduces the need to improve the increase in yield in quantitative and qualitative terms, more effective crop protection methods and more efficient management systems. What's more, yield forecasting is an important tool used for extreme weather conditions as they are becoming more and more frequent and their effects are negative and long-lasting.

2. Historical approach to plant yield prediction

The history of methods used to predict crop yields dates back to the beginning of the 20th century. One of the methods was taking average of three, which assumes that the actual yield to be achieved in a given year is the arithmetic average of the three previous growing seasons. Unfortunately, there are very few years in which this approach can give a proper forecast. For instance, when a particular season that was favourable for growth and development of plants, and the previous three were not very successful. In such cases, the prediction error can reach up to several hundred percent. Another old method, called fertilizer method, combines yielding of plants with the amount of used fertilizers. Fertilizer unit - the basic predicting parameter, expresses 3 kg of mineral fertilizers calculated as the pure component in the NPK ratio: 1: 0,8: 1. Observations from the 50s of the last century allowed to conclude that one fertilizing unit allows for the production of 20 kg of grain [31]. The best effects in yielding plants are obtained while maintaining the above-mentioned proportion of mineral components by determining the fertilization within the farm.

The presented methods, despite their few advantages, do not take into account one of the most important factors that determine crop yielding, such as meteorological conditions. Due to the random nature, often limiting the yield potential of plants, the weather conditions are usually treated as distortions [27]. In the method proposed by Zalewski, the yield is predicted by a linear relation between the consumption of mineral fertilizers and weather conditions. The climate, as a complex factor, was presented in the form of a hydrological and thermal quotient, so-called Sielianinowa coefficient [16]. The author of the method states that the following coefficient should be calculated in detail for the month of May due to the experimentally confirmed relationship between final yields and weather conditions. It turned out, however, that the actual yields achieved in subsequent years of research compared to the yields calculated on the basis of the discussed method did not differ significantly. Another approach to yield prediction may be a detailed observation of plants in critical and difficult development phases. In case of potatoes, the setting period for stolons was adopted [31]. At this stage, the influence of all factors significantly influencing the yielding is seen, and the intensification of specific symptoms can initially suggest a level of crop yield.

3. Current trends in yield prediction

The current approach in estimating plant yield has evolved along with the development of measuring methods and techniques. The most common solutions include ground-based remote sensing using models simulating plant development or classical regression models, artificial intelligence methods, image analysis.

3.1. Remote sensing and vegetation indices

Remote sensing methods allow for observation and control of crops from air and satellite via special satellite images, e.g. NOAA AVHRR used by the Institute of Geodesy and Cartography that create a system for monitoring the state of arable crops in Poland. As for yield prediction, the AVHRR NOAA scanners, present on the satellites, register the radiation on the basis of which the NDVI index values (Normalized Difference Vegetation Index) are determined. The index developed by Rouse and others [29] is a quotient of difference and sum of radiation reflected from the surface of plants in the near infrared (IR) and infrared (R). Special compositions and maps are created on the basis of daily images of index distribution over areas of agriculture use. Also, the analysis of changes in the index value in the growing season and multiannual period is performed. Numerous studies indicate a visible, high correlation of vegetation indices in respect to yield and plant biomass [22]. Plant life studies have shown that NDVI is associated with the leaf area index (LAI) and photosynthetic crop activity. The effect of conducting the following research consists in developing a model (usually based on regression technique) - an empirical relation of values of vegetation indices measured during the vegetative cycle (Fig. 1) and the yield of specific species. A small output database and ease of making forecasts is the big advantage of such models. Creating models based on measurements carried out in strictly defined environmental conditions may bring certain difficulties concerning their universal application, i.e. high uncertainty of a reliable forecast on a wider scale [26]. However, the use of NDVI data has limitations, i.e. calibration of the device, incidence of solar radiation, atmospheric attenuation and cloud occurrence, which limits the use of the model [3].

Source: picture by / Źródło: fot. K. Piekutowski

Fig. 1. The Crop Circle OptRx®- an NDVI sensor
Rys. 1. Czujnik Crop Circle OptRx® do pomiaru NDVI

The use of remote sensing enabled the model to predict the yield of irrigated potato plantations by the end of the growing season using vegetation indices NDVI, SAVI, CSAVI and CNDVI. Potato yield samples were collected 2-3 days before harvest and were correlated with neighboring NDVI and SAVI, where yield prediction algorithms were developed and used to generate yield prediction maps. Relatively low prediction errors were observed, amounting to no more than 10% [2]. The relationship between the actual and predicted yield value, expressed in the R2 coefficient of determination, was 0,65. This value indicates a satisfactory model match. According to the authors, the period of 60-70 days after planting potatoes is the best time to make reliable predictions.

3.2. Classic forecasting models

To assess the impact of input variables on potato yield, authors often use the classic approach, like the multiple regression analysis method [14], stepwise regression [15]. Regression models are a kind of formal record of dependencies between dependent variables, e.g. yield size and independent factors that shape this feature in the form of a mathematical equation. In other words, the regression calculation problem is to match a straight line to a set of points. The value of the determination coefficient R2 ranges from 0 to 1, which tells us what part of the dependent variable was described by the model. There are some research results which prove that traditional methods of plant yield prediction may be characterized by low prediction accuracy. Classical models are linear, and in many cases this approximation of dependencies is incorrect. Moreover, these methods usually depend on strict field data collection on individual yield and yields in general, which requires a lot of costs and time. Models for plant yield forecasting can be built based on a variety of sets of features and measurements. The important part is that the parameters chosen should actually shape the value of the forecasted variables. The agrometeorological data typical for a given growing season are the most universal and useful data, taken into account in the construction of yield prediction models [14]. Most models created over the years for potato crops include such information. NPK fertilization doses are another important and popular data [4].

3.3. Basic potato models

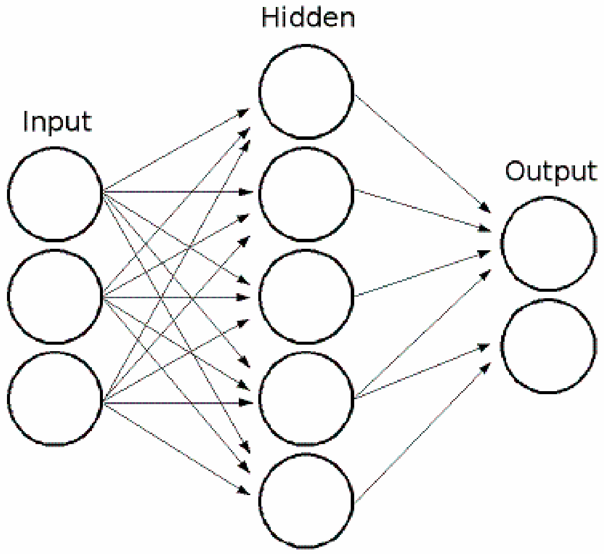
Currently, the most popular models in the world are: SUBSTOR [28] and LINTUL-POTATO [20]. SUBSTOR (Simulate Underground Bulking Storage Organs) was built based on the common cereal model CERES [13]. It was designed as a decision supporting system for potatoes, acting as a tool to explain plant-substance relations. SUBSTOR is widely available and used in many countries worldwide. The most important applications include: determining the optimal planting time, controlling the irrigation of plantations, optimizing nitrogen fertilization doses. The model simulates plant development and agronomic yields depending on weather conditions, soil and cultivation. LINTUL-POTATO (Light INTerception and utilization for POTATO) was created in the Netherlands basing on studies that deal with the influence of temperature and day length on the growth and development of eight potato varieties. Currently, it is being applied in South America, as the agro-ecological characteristics of potato - identifying factors determining yield, limiting and reducing yield, the possibility and profitability of transferring new agricultural methods, e.g. new varieties.

3.4. Artificial neural networks in yield prediction

Artificial neural networks are an increasingly used tool that supports research in agriculture. The main advantages include large approximation abilities as well as many independent variables in the analysis (also of linguistic nature) solving non-linear dependencies. The idea of building and functioning of artificial neural networks derives from the observation and imitation of the human brain and nervous system. Artificial Neural Network ANN is a way of defining mathematical models and structures which take part in complicated calculations. Their application is very wide, as they match any model whose degree of complexity of the tested dependent and independent variables is high and not possible to be expressed through classical correlation. ANNs are able to map non-linear dependencies [12] and their great advantage consists in their self-learning capability. Based on the created database containing all the relevant variables and the learning algorithm, the network remembers the data structures. Thanks to self-created weighting factors, the network performs all further tasks, which ultimately leads to forming the model [9]. The neural network is characterized by architecture, a training algorithm and an activation function. The network's architecture is understood as the number of layers forming it: input, output, one or two hidden layers [8]. Network learning means "forcing" a specific response by means of specific signals. The network can learn to perform calculations in two ways, i.e. by supervised learning or previously mentioned self-learning. The activation function is a tool for calculating the value of neuron output [11]. Moreover, neural networks have been found useful in yield prediction of plants [5], in operation and automation systems [10], classifying and predicting many diseases [17], optimization and classification in business [21]. Artificial intelligence methods become more and more important in the aspect of potato yielding, being used for all types of predictions: environmental indices of potato production [18], virescence of tubers [6], profitability of production [35], energy consumption of agricultural machinery [34]. The parameters defining the qualitative characteristics of the created models indicate the predominance of neural models over other predictive models. The MAPE error value (mean absolute percentage error) - the basic qualitative indicator of the models, often does not exceed 10-15%. Such values show the adequate knowledge of the mechanisms and dependencies occurring in the production of potatoes and processes in plants.

Artificial neural networks are a perfect tool for forecasting yields of potatoes, as they allow for accurate analysis of a dynamic process depending on many random factors. Different types of networks are used to predict yields. Two types of networks were tested for Radial Basis Function Neural Network (RBFNN) and General Regression Neural Network (GRNN) in two combinations, the Flat Potato Field (FPFs) and Rough Potato Field (RPFs) by Pandey and Mishra [25]. RBF networks consist of three layers. The input layer allows you to connect the network to the environment. The next layer - hidden, consists of several nodes that use non-linear transformation to input variables, using a radial basic function, such as the Gauss function. The output layer is linear and is called the summing unit [30]. The GRNN network structure consists of four layers: the input layer, two hidden layers and the output layer. The first hidden layer of the network consists of radial neurons. The second layer consists of only two neurons, which estimate the weighted average of the neuron outputs of the previous layer. The first neuron marks the so-called weighted sum. The second neuron is responsible for the sum of the weighting factors. Obtaining a weighted average from a weighted sum is possible by the action of a neuron in the hidden layer, whose task is to make calculations: quotient of the sum weighted by the sum of weighting factors. This task is performed by the neuron in the output layer [32]. According to some literature sources, the GRNN network has a greater learning ability with a large amount of data compared to RBFN [19].

In the Pandey and Mishra research [25], the GRNN network proved to be a more accurate predictor of potato yield based on three yield parameters: average plant height, leaf area index (LAI), stem and leaf dry weight than RBNN. Fortin et al. [8] used another type of neural network - MLP Multiple-layer perceptron (Fig. 2) to predict potato yield. A characteristic feature of the MLP network is connected with the presence of at least one layer of hidden neurons, which mediates signals transfer between the input nodes and the output layer. Prediction of tuber yield was based on various combinations of input variables: LAI values, meteorological data including solar radiation and such data as: the following day of the year.



Source: / Źródło: https://docs.opencv.org/3.4/dc/dd6/ml_intro.html

Fig. 2. General structure of the MLP artificial neural network

Rys. 2. Struktura sztucznej sieci neuronowej o topologii MLP

The most influential variables on the yield of tubers were solar radiation and the sum of precipitation. The use of MLP networks with three independent variables: sum of rainfall, radiation and LAI values allowed for the most accurate simulation of yield increase during the growing season. The MLP network with two layers was also used to predict the starch content in potato tubers. At the input of network, the authors included as many as 14 independent variables, and 56 varieties of potatoes from different habitat conditions were analyzed. The generated model was characterized by high quality and can create the basis for the functioning of the decision supporting system in starch content forecasting [24].

3.5. Practical use of image analysis methods

From the agricultural point of view, it's a huge success to transfer works from the preliminary analysis level of research teams to specific tools offered to agricultural producers. The decision supporting products in agriculture and everyday applications can be of such importance. The SolGrader application created by Solentum company from the Netherlands is a good example of knowledge transfer in the field of forecasting potato yields. It is enough for the manufacturer to complete the program with basic cultivation data: date of planting, area, planting distance, row width. The prediction is based on image analysis.

The parameters to be measured are as follows: the size and mass of tubers, the average length of tuber, the estimated total yield and the initial calibration. The recipients of the application are potato producers who can more precisely and professionally decide on the date of desiccation and lifting by following the growth of tubers, as well as determine the necessary storage area for crops. Thanks to the application, food processing plants will make initial and quick decisions about the order of raw material deliveries, for example during uncommon years for potato production, when it is necessary to adjust the potato delivery plan due to their difference in quality [33]. Image analysis is a tool that allows not only for the quantitative prediction of yields or raw materials, but also the prediction of product quality possible to obtain in industrial conditions [23].

4. Summary

Prediction of agricultural crop yield is a modern approach in effective management of farm. The current trend of sustainable agriculture, being a less extreme form of the intensive system, indicates the need for wider use of modern technologies to realize maximum profits while caring for the environment. Thanks to the access to tools that implement predictive tasks on a daily basis, producers can not only forecast yields, but also decide beforehand about possible deficiency in production resources depending on the conditions in the growing season.

5. References

[1] Al-Gaadi K.A., Hassaballa A.A., Tola E., Kayad A.G., Madugundu R., Alblewi B., Assiri F.: Prediction of potato crop yield using precision agriculture techniques. Plos One, 2016, 11.

[2] Baez-Gonzalez A.D., Kiniry J.R., Maas S.J., Tiscareno M.L., Macias J.C., Mendoza J.L., Richardson C.W., Salinas J.G., Manjarrez J.R.: Large-area maize yield forecasting using leaf area index based yield model. Agronomy Journal, 2005, 97, 418-425.

[3] Benedetti R., Rossini P.: On the use of NDVI profiles as a tool for agricultural statistics: The case study of wheat yield estimate and forecast in Emilia Romagna. Remote Sensing of Environment, 1993, 45, 311-326.

[4] Cerrato M.E., Blackmer A.M.: comparison of models for describing; corn yield response to nitrogen fertilizer. Agronomy Journal, 1988, Vol. 82 (1), 138-143.

[5] Dahikar M.S.S., Rode S.V.: Agricultural crop yield prediction using artificial neural network approach. International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering, 2014, Vol. 2 (1), 1-4.

[6] Edwards E.J., Cobb A.H.: Effect of Temperature on glycoalkaloid and chlorophyll accumulation in potatoes (*Solanum tuberosum* L. Cv. King Edward) stored at low photon flux density, including preliminary modeling using an Artificial Neural Network. J. Agric. Food Chem., 1997, Vol. 45 (4), 1032-1038.

[7] Ericson G.R.: A comparison of models to forecast annual average potato prices in Utah. Utah State University, 1993.

[8] Fortin J.G., Parent L.E., Anctil F., Bolinder M.A.: Neural network to simulate potato tuber yield in eastern Canada. Acta horticulturae, 2008, Vol. 802 (802), 309-318.

[9] Graupe D.: Principles of Artificial Neural Networks (2nd Edition). Advanced Series on Circuits and Systems, 2007, 6.

[10] Hagan M.T., Demuth H.B., Jesús O.D.: An introduction to the use of neural networks in control systems. Int. J. Robust Nonlinear Control, 2002, Vol. 12 (11), 959-985.

[11] Jain A.K., Mao J., Mohiuddin K.M.: Artificial Neural Networks: a tutorial. Computer, 1996, Vol. 29 (3), 31-44.

[12] Jeż R.: Sztuczne sieci neuronowe i ich zastosowanie w modelowaniu zjawisk gospodarczych. Zeszyty Naukowe Uniwersytetu Ekonomicznego w Katowicach, 2018, 217, 182-191.

[13] Jones C.A., Kiniry J.R.: Ceres-maize: A simulation model of maize growth and development, Texas A & M University Press, College Station, Texas, 1986.

[14] Kalbarczyk R.: An attempt of potato yield forecasting on the basis of agrometeorological data in Poland. Folia Univ. Agric. Stetin. Agricultura, 2004, Vol. 234 (93), 145-158.

[15] Kalbarczyk R.: Czynniki agrometeorologiczne a plony ziemniaka w różnych rejonach Polski. Acta Agrophysica, 2004, Vol. 4 (2), 339-350.

[16] Kapuściński J., Nowak R.: The frequency of the occurrence of droughts and post-droughts periods in mid-west Poland on the example of Poznań, Wałcz and Wieluń. Management and conservation of the forest environment. Climate vs forest (Ed. Miler A.) Agricultural University in Poznań, 2003, 76-88.
[17] Khan J., Wei J.S., Ringner M., Saal M.H., Ladanyi M., Westermann F., Berthold F., Schwab M., Antonescu C.R., Petereson C., Meltzer P.S.: Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. Nature Med. 2001, 7, 673– 679.
[18] Khoshnevisan B., Rafiee S., Omid M., Mousazadeh H., Sefeedpari P.: Prognostication of environmental indices in potato production using Artificial Neural Networks. Journal of cleaner production, 2013, 52, 402-409.
[19] Kim B., Kim S., Kim K.: Modeling of plasma etching using a generalized regression neural network. Vacuum, 2003, 71, 497-503.
[20] Kooman P. L., Haverkort A.J.: Modelling development and growth of the potato crop influenced by temperature and daylength: LINTUL-POTATO In Potato ecology and modelling of crops under conditions limiting growth (Eds, Haverkort A.J. and D.K L. Mackerron) Kluwer Academic Publisher Netherlands, 1995, 41-60.
[21] Li E.Y.: Artificial neural networks and their business applications. Information & Management, 1994, 27, 303-313.
[22] Liu W.T., Kogan F.: Monitoring Brazilian soybean production using NOAA/AVHRR based vegetation condition indices. International Journal of Remote Sensing, 2002 Vol. 23(6), 1161-1179.
[23] Marique T., Kharoubi P., Bauffe P., Ducatillon C.: Modelling of fried potato chips color classification using image analysis and Artificial Neural Network. Journal of Food Science, 2003, 68, 2263-2266.
[24] Niedbała G., Lenartowicz T., Kozłowski R., Zaborowicz M.: Modelowanie neuronowe jako metoda prognozowania zawartości skrobi w ziemniakach na potrzeby Porejestrowego Doświadczalnictwa Odmianowego i Rolniczego (PODiR). Nauka Przyroda Technologie, 2015, Vol. 9 (2).
[25] Pandey A., Mishra A.: Application of Artificial Neural Networks in yield prediction of potato crop. Russian Agricultural Sciences, 2017, Vol. 43 (3), 266-272.
[26] Piekarczyk J.: Szacowanie plonów roślin uprawnych na podstawie naziemnych pomiarów spektralnych. Teledetekcja środowiska, 2011, 46, 23-28.
[27] Powell J.P., Reinhart S.: Measuring the effects of extreme weather events on yields. Wheater and Climate Extremes, 2016, 12, 69-79.
[28] Ritchie J.T., Griffin T.S., Johnson B.S.: SUBSTOR: Functional model of potato growth, development and yield In Modelling and parameterization of the soi/-plant-atmosphere system: A comparison of potato growth models. (Eds, Kabat P.B., Marshall B., van den Broek J., Vos H. van Keulen) Wageningen Pers, Wageningen Netherlands, 1995.
[29] Rouse J.W., Haas R.H., Schell J.A., Deering D.W.: 1973. Monitoring vegetation systems in the Great Plains with ERTS. Conference Paper of NASA. Goddard Space Flight Center 3d ERTS-1 Symp, 1976, Vol. 1, Sect. A, 309-317.
[30] Sarimveis H., Alexandridis A., Mazarakis S., Bafas G.: A new algorithm for developing dynamic radial basis function neural network models based on genetic algorithms. Computer Aided Chemical Engineering, 2002, 10, 949-954.
[31] Sobczyk M.: Zagadnienie prognozowania w świetle literatury. Annales UMCS, Sectio H., 1976, Vol. 10 (12), 183-200.
[32] Sobolewski Ł.: Comparison of the usefulness of GRNN and RBF neural networks for predicting the corrections for the national time scale UTC(PL). Pomiary Automatyka Kontrola, 2014, 60, 972-974.
[33] SolGrader-Solentum, https://solentum.com/product/solmat-a2/.
[34] Zangeneh M., Omid M., Akram A.: Assessment of machinery energy ratio in potato production by means of Artificial Neural Network. Africal journal of agricultural research, 2010, Vol. 5 (10), 993-998.
[35] Zangeneh M., Omid M., Akram A.: A comparative study between parametric and artificial neural networks approaches for economical assessment of potato production in Iran. Spanish journal of agricultural research, 2011, 3, 661-671.